Lesson 2\_1

**import all libraries**

import warnings

warnings.filterwarnings("ignore")

import matplotlib.pyplot as plt

import seaborn as sns

%config InlineBackend.figure\_format='retina'

plt.rcParams["figure.figsize"] = 8, 5

plt.rcParams["image.cmap"] = "viridis"

import pandas as pd

**read dataset**

import pandas as pd

df = pd.read\_csv("video\_games\_sales.csv").dropna()

print(df.shape)

**display information**

df.info()

**convert object to float and int datatype**

df["User\_Score"] = df["User\_Score"].astype("float64")

df["Year\_of\_Release"] = df["Year\_of\_Release"].astype("int64")

df["User\_Count"] = df["User\_Count"].astype("int64")

df["Critic\_Count"] = df["Critic\_Count"].astype("int64")

**create useful columns**

useful\_cols = ["Name", "Platform", "Year\_of\_Release", "Genre", "Global\_Sales", "Critic\_Score", "Critic\_Count", "User\_Score", "User\_Count", "Rating"]

df[useful\_cols].head()

**plotting**

df[[x for x in df.columns if "Sales" in x] + ['Year\_of\_Release']].groupby('Year\_of\_Release').sum().plot()

**plotting chart bar**

df[[x for x in df.columns if "Sales" in x] + ["Year\_of\_Release"]].groupby("Year\_of\_Release").sum().plot(kind="bar", rot=45)

**pairplotting chart bar**

import seaborn as sns

sns.pairplot(df[["Global\_Sales", "Critic\_Score", "Critic\_Count", "User\_Score", "User\_Count"]])

**displot chart bar**

%config InlineBackend.figure\_format = 'retina'

sns.distplot(df["Critic\_Count"])

**joinplot chart bar**

sns.jointplot(x="Critic\_Score", y="User\_Score", data=df, kind="scatter")

**boxplot chart bar**

top\_platforms = (df["Platform"].value\_counts().sort\_values(ascending=False).head(5).index.values)

sns.boxplot(

y="Platform",

x="Critic\_Score",

data=df[df["Platform"].isin(top\_platforms)],

orient='h'

)

**Heatmap chart bar**

platform\_genre\_sales = (

df.pivot\_table(

index="Platform", columns="Genre", values="Global\_Sales", aggfunc=sum

)

.fillna(0)

.applymap(float)

)

sns.heatmap(platform\_genre\_sales, annot=True, fmt=".1f", linewidths=0.5)

**plotly**

import plotly

import plotly.graph\_objs as go

from plotly.offline import download\_plotlyjs, init\_notebook\_mode, iplot, plot

init\_notebook\_mode(connected=True)

**line plot**

years\_df = (df.groupby("Year\_of\_Release")[["Global\_Sales"]].sum().join(df.groupby("Year\_of\_Release")[["Name"]].count()))

years\_df.columns = ["Global\_Sales", "Number\_of\_Games"]

**trace chart bar**

trace0 = go.Scatter(x=years\_df.index, y=years\_df["Global\_Sales"], name="Global Sales")

trace1 = go.Scatter(x=years\_df.index, y=years\_df["Number\_of\_Games"], name="Number of games released")

data = [trace0, trace1]

layout = {"title": "Statistics of video games"}

fig = go.Figure(data=data, layout=layout)

iplot(fig, show\_link=False)

**bar chart**

platforms\_df = (

df.groupby("Platform")[["Global\_Sales"]].sum().join(df.groupby("Platform")[["Name"]].count())

)

platforms\_df.columns = ["Global\_Sales", "Number\_of\_Games"]

platforms\_df.sort\_values("Global\_Sales", ascending=False, inplace=True)

another trace chart bar

trace0 = go.Scatter(x=years\_df.index, y=years\_df["Global\_Sales"], name="Global Sales")

trace1 = go.Scatter(x=years\_df.index, y=years\_df["Number\_of\_Games"], name="Number of games released")

data = [trace0, trace1]

layout = {"title": "Market share by gaming platform"}

fig = go.Figure(data=data, layout=layout)

iplot(fig, show\_link=False)

**another box plot**

data = []

for genre in df.Genre.unique():

data.append(go.Box(y=df[df.Genre == genre].Critic\_Score, name=genre))

iplot(data, show\_link=False)